

# GIS中矢量栅格一体化数据结构的研究\*

龚 健 雅

(武汉测绘科技大学)

## 提 要

数据结构在地理信息系统(GIS)中占有突出的地位,它直接关系到GIS的功能和效率。新一代地理信息系统要求能够统一管理GIS、DEM和RS的各种数据。本文提出一种既有矢量特点又有栅格性质的一体化数据结构用来支持这三者结合的系统。这种矢量栅格一体化的数据结构以线性四叉树编码为基础,采用多级格网方法以提高GIS目标栅格表达的几何精度和快速索引。它能代替传统的矢量和栅格结构,并且可以保持较好的精度,压缩存储空间,表达各种拓扑关系,支持面向目标的数据模型以及整体空间数据库。

关键词 地理信息系统 线性四叉树编码 一体化数据结构 面向目标数据模型

## 1 引 言

传统的GIS一般是基于矢量的系统,这种系统具有精度高、容易操作目标等优点,但它与DEM和遥感数据(RS)交互等方面的能力较差。在GIS建立以后,遥感数据将成为更新GIS和评价地理环境等方面的主要信息源。适应于GIS与DEM和RS直接交互的系统只能是基于栅格的系统。但栅格数据分辨率低、数据量大,不便直接定义单个目标和拓扑关系,使单纯基于栅格的系统难以普及应用。

地理信息系统与遥感分析系统的结合已成为二者发展的必然趋势。GIS与RS的结合存在三种可能的方式。<sup>[1]</sup>第一种称为“分开但是平行的结合”,它是由一个传统的GIS联系到一个遥感图像处理系统,两个系统的用户界面是分开的,这是目前大部分系统的结合方式。第二种方式称为“无缝的结合”,从外部看,它们处于同一用户环境中,但图像的栅格数据与GIS中的空间数据仍分开存放,这类系统目前正处于研究开发之中。理论上,专家们正在探讨第三种水平的结合,即完全结合的系统,这种系统不仅使用同一用户界面,而且属性、矢量和栅格数据(包括DEM)存放在一个数据库中,利用统一的DBMS管理。在这种综合的GIS中无需变换就可以对各种栅格数据和矢量数据进行查询、处理和显示。本文提出一种一体化的数据结构同时表达矢量和栅格的特性以支持GIS与DEM和RS完全结合的系统。

## 2 二维行程编码和多级格网技术

在栅格数据压缩方面,通常使用行程编码法和四叉树方法。四叉树方法过去常被认为算法复杂,增删数据困难、应变性差而难以被人接受。<sup>[4]</sup>这里介绍一种兼有两者特性的编码方案

\* 本文1991年8月13日收到,截稿日期为1992年4月10日。

本文是博士论文的一部分,指导教师为王之卓教授、李德仁教授和Ole Jacob教授

——二维行程编码，可以克服四叉树的某些缺点，并提高栅格数据的压缩效率。

### 2.1 线性四叉树与二维行程编码

目前人们一般采用线性四叉树取代常规四叉树管理栅格数据，而且采用Morton地址码表示四叉树叶结点的位置。Morton码常用的有四进制码和自然数的十进制码，文献[5]已经论证十进制码使用更为方便。按照图1(a)所示的Morton码排列成线性表图1(b)，用自底向上的合并方法可产生四叉树的线性表图1(c)。

我们注意到，在生成的线性四叉树表列图1(c)中，仍存在前后叶结点的值相同的情况，因而采取进一步的压缩表达方法，即将灰度值相同的前后叶结点合并成一个值，形成图1(d)所示的线性表列。图1(d)也可以从图1(b)中直接得到，先记录入口地址和灰度值，依次序扫描图1(b)的线性表，若后一格网的灰度值不等于前一格网的灰度值，记录后一格网的Morton地址码和相应的灰度值，直接可形成图1(d)的线性表。这种记录编码的方法非常类似于传统的行程编码，所以称它为二维行程编码。在这种二维行程编码中，前后两个地址码之差表达了该游程段的格网数，它可以表达该区域子块的大小。

这种二维行程编码利用了线性四叉树的地址码，但是没有构造规则的四叉树，甚至已失去了四叉树的概念，然而它比规则的四叉树更节省存储空间，而且对于以后的删除和修改，因不必保持完整的四叉树构形而变得相当简便。

当然，线性四叉树本身具有许多良好的特性，叶结点之间有良好的图形拓扑关系，<sup>(6)</sup>可对它们进行图形几何分析，<sup>(5)</sup>以及快速区域填充和栅格矢量化。<sup>(7)</sup>由于线性四叉树和二维行程编码采用相同的地址码，它们之间的相互转换非常容易和快速，在我们讨论的数据结构中几乎将它们视为同一概念，只是根据需要采用不同的形式。

### 2.2 多级格网技术

目前人们之所以无法丢弃矢量数据的重要原因，是因为栅格数据的精度太低。这里提出一种细分格网方案来提高点、线的表达精度。如图2所示，在有边界通过的基本格网内，再细分成256×256个细格网(精度要求低些时，可细分成16×16个细格网)，同样采用线性四叉树的编码方法，将采样点以及线性目标与基本格网边的交点用两个Morton码表示。前一个码M<sub>1</sub>表示

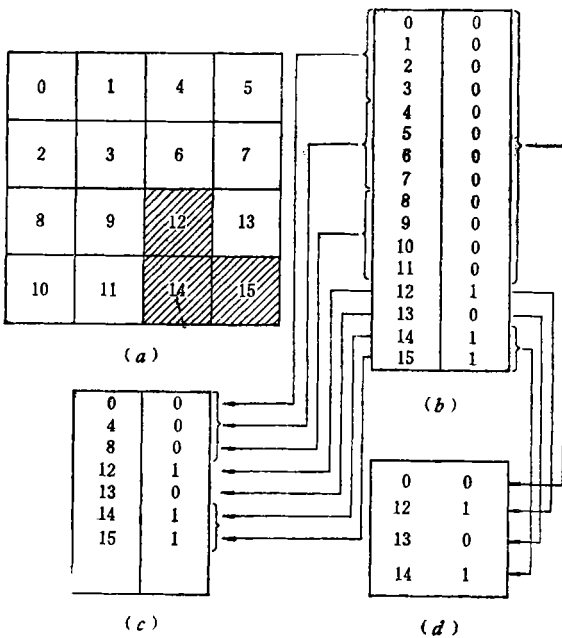


图1 线性四叉树与二维行程编码过程

示该点(采样点或附加的交叉点)所在的基本格网的地址码。后者 M<sub>2</sub> 表示该点对应的 Morton 码。即将对 x, y 坐标转换成两个 Morton 码。

例如当栅格间距为10,

$$x=210\ 00 \quad y=172\ 32$$

可以转换成:

$$M_1=275 \quad M_2=2690$$

这种方法可将栅格数据的表达精度提高256倍,而存储量仅在有点、线通过的格网上增加2bytes(当细分成 $16 \times 16$ 个细格网时,存储量仅增加1byte,精度提高16倍)。

另一方面,再采用粗格网方法建立索引文件。目前对于一般的大型数据库通常采用B树或B<sup>+</sup>树方法索引。但GIS中的几何数据,DEM和遥感数据是空间数据,它们与位置有密切的关系,所以采用线性四叉树索引方法更为有效,它使索引记录本身隐含位置信息。

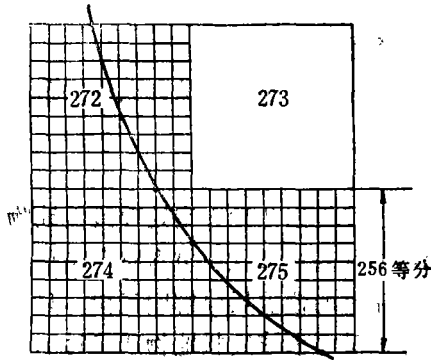


图2 细分格网

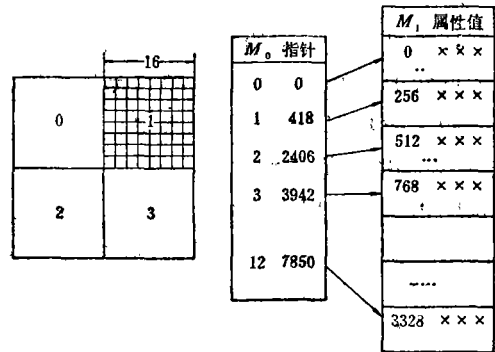


图3 线性四叉树的粗网格索引

将 $16 \times 16$ 个基本格网组成一个粗格网,每个粗格网也用十进制的Morton编码。这些粗格网组成一个索引表,显然它的顺序是按线性四叉树的地址码排列的。索引表中每个记录用一个指针,指向对应粗格网包含的第一个基本格网在线性四叉树表或二维行程编码表中的位置。为了统一起见,每个四叉树的叶结点或二维行程编码的游程段不能超过一个粗格网。这样,便建立了索引记录与线性表的关系(如图3)。

根据某点的位置可直接进入索引记录,进而找到该点在线性四叉树表中的位置。例如,查找 $M_1=4235$ 的块属性值,先求它的粗格网码 $M_0=4235/256=12$ ,直接进入索引文件中的第13号记录(记录号= $M_0+1$ ),顺着指针找到该粗格网在线性四叉树表中的起始记录号,再往下查找,即可得到该叶结点的属性值。注意到这里不需要搜索索引表,而是直接进入13号记录。另外它还可以省去索引文件中的Morton码,用隐含的记录号代替。

### 3 GIS中的一体化数据结构

线性四叉树编码和多级格网技术为建立矢量与栅格一体化的数据结构奠定了基础。在讨论数据结构之前,先强调三个假设:

A 地面上的点状地物是地表上的点,它仅有空间位置,没有形状和面积。

B 地面上的线性地物是地球表面上的空间曲线,它有形状,但没有面积,它在平面上的投影是一连续不间断的直线或曲线。

C 地面上的面状地物是地表面的空间曲面,并且有形状和面积,它在平面上的投影是由边界和边界包围的一个无孔紧致空间构成。

这三个基本假设将指导我们建立空间数据的结构。

#### 3.1 点状地物和结点的数据结构

根据假设A,点仅有位置,没有形状和面积,不必将点状地物作为一个覆盖层分解四叉树,只需将点的坐标直接转化为前面所述的Morton码 $M_1$ 和 $M_2$ ,而无需考虑整个结构是否为四

表 1 点状地物的数据结构

点 标 识 号	$M_1$	$M_2$	高 程
...	...	...	...
10025	43	4082	432
10026	105	7725	463

叉树。这样既可保证一定的精度，又允许一个格网内包含多个点状地物。

这种点状地物的数据结构几乎与矢量数据的结构完全一致。它比文献〔8〕、〔9〕所述的点状地物的四叉树数据结构要简单得多。

### 3 2 线状地物的数据结构

线状地物的四叉树表达被认为是最困难的。<sup>〔9〕</sup>目前已经提出几种方案处理线性地物，但都相当复杂。根据假设 B，线性地物有形状没有面积，而且表达形状应包含整个路径。没有面积这就说明，线状地物与点状地物一样，不必用一个完全的覆盖层分解四叉树，只要用一串数字表达每个线状地物的路径即可。要表达整个路径，又要求该线状地物所经过的格网要全部记录下来。

一个线状地物可能由  $n$  条弧段组成，所以先建立一个弧段的数据文件如表 2 所示。

表 2 弧段的数据结构

弧 标 识 号	起 结 点 号	终 结 点 号	中间点串( $M_1, M_2, Z$ )
...	...	...	.....
20078	10025	10026	58, 77, 49, 435, 92, 4377, 439
20079	10026	10027	105, 432, 502, 112, 44, 2, 4%

表中起结点和终结点是该弧段的两个端点，它们与表 1 联接可建立弧段与结点之间的拓扑关系。表 2 中的中间点串不仅包含了原始取样点，而且包含了该弧段路径通过的所有格网边的交点，中间点所包含的  $M_i$  填满了整个路径。这样的结构也充分考虑了线性地物在地表的空间特性。若一条线状地物是在崎岖的地面上通过的，只有记录曲线通过 DEM 格网边上的高程值才可能较好地表达它的空间形状和长度。

这种数据结构比单纯的矢量结构增加了一定的存贮量，但它解决了线性地物的四叉树表达问题，使它能与点状地物和面状地物一起建立统一的基于线性四叉树编码的数据结构体系，对于点状地物与线状地物的相交，线状地物之间的相交，以及线状地物与面状地物相交的查询问题变得相当简便和快速。<sup>〔7〕</sup>

有了弧段数据文件，线性地物的数据结构仅是弧段的集合表示，如表 3 所示。

表 3 线状地物的数据结构

线 标 识 号	弧 标 识 号 串
...	...
30031	20078 20079
30032	20092 20098 20099

### 3.3 面状地物的数据结构

按照基本假设 C, 一个面状地物应包含边界和边界所包围的整个面积。面状地物的边界也是由弧段组成的, 它同样引用表 2 中的弧段信息。除此以外, 它还要包含面域的信息。面域信息则由二维行程编码表示。

各类面状地物可能形成多个覆盖层。例如地面上所有的可视地物可作为一个覆盖层, 行政区划和土壤类型又可形成另外两个覆盖层。这里规定每个覆盖层都是单值的, 即每个栅格内仅有一个面状地物的属性值。一个覆盖层是一个紧致空间, 即使是孤岛也含有相应的属性。每个覆盖层要用一个二维行程编码的线性表表示, 其结构如表 4。

表 4 二维行程编码线性表

基本格网 $M_1$ 码	指针
0	8
32	68
66	75
68	143
75	99
88	105
99	40002

线性表是按基本格网的 Morton 码顺序排列的, 表 4 中第二列的指针代替了图 1(d) 中的属性值, 其目的是为了建立面向目标的数据模型。这种循环指针指向该地物的下一个子块的记录, 并在最后指向该地物本身, 只要进入第一块就可以直接提取该地物的所有子块。这样可以避免像栅格矩阵那样, 为了查询某一个目标而搜索整个矩阵, 从而可大大加快查询速度。

对于面状地物的边界格网, 采用以面积为指标类似四舍五入的方法确定它的属性, 这就是说两地物的公共格网其属性取决于该格网内面积比重大的那个地物。如果要精确进行面状地物的面积计算和迭加运算, 可以进一步引用弧段的边界信息参与计算或修正。

表 2 和表 4 是面状地物数据结构的基础, 它的具体形式见表 5。

表 5 面状地物的数据结构

面标识号	弧标识号串	面块头指针
40001	20001, 20002,	1
40002	20002 20004,	3

表 6 复杂地物的数据结构

复杂地物标识号	简单地物标识号串
50008	10025, 30008, 30025
50009	30006, 30007, 40032

表 5 中的头指针指向二维行程编码表中该地物对应的第一个子块。

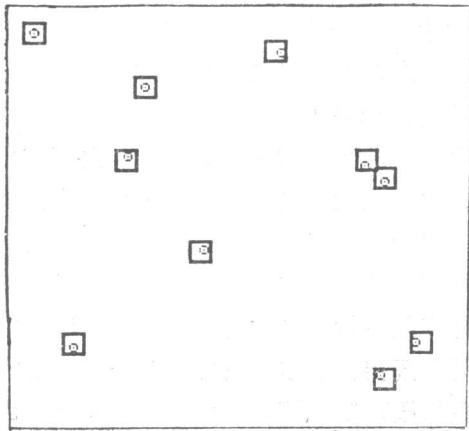
这种结构是面向目标的, 并具有矢量的特点, 只要找到面块标识号, 就可以通过指针找出它的边界弧和中间面块。同时它又具有栅格的全部特征。二维行程编码表本身就是面向位置的结构。表 4 中的 Morton 码表达了位置的相关关系, 前后两个 Morton 码之差隐含了该子块的大小。一个覆盖层形成了一个二维行程编码表, 从第一个记录到最后一个记录表示的面块覆盖了工作区域的整个平面, 因而满足了基本假设 C。给出任一点的位置都可在二维行程编码表中顺着指针找到弧和面状地物的标识号确定是哪一个目标。

### 3.4 复杂地物的数据结构

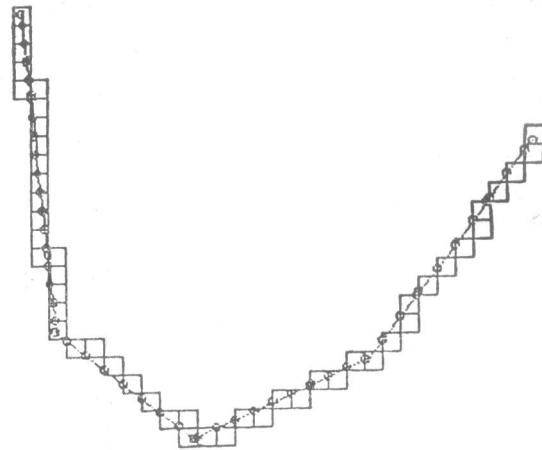
单独用一个点, 或一条线, 或一个面状地物组成的地物称为简单地物。由  $n$  个或  $n$  种简单地物组成的地物称为复杂地物。组合复杂地物的概念在面向目标的数据模型中起了很重要的作用。例如我们可以将一条公路上的中心线、公路面、交通灯及立交桥等作为一个复杂地物, 用一个标识号表示。复杂地物的数据结构如表 6。

注意表 6 右边的简单地物标识号中的第一位数不完全相同, 说明它可以由不同类型的地物组成。而表 3 和表 5 中的标识号的第一位数总是相同的。

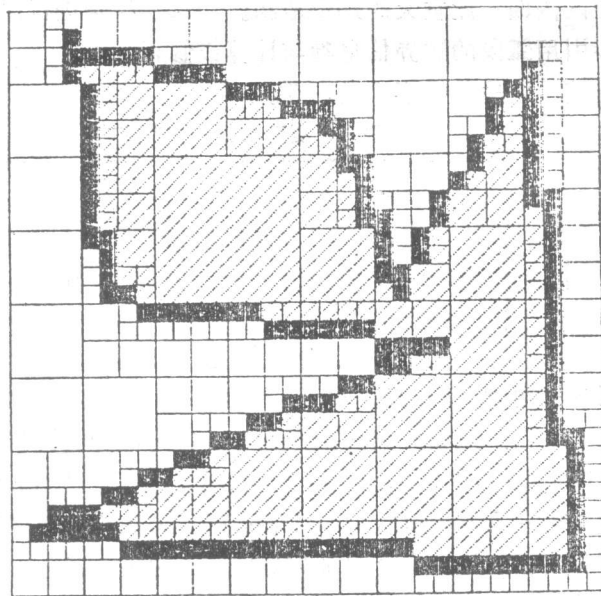
根据以上讨论的四类地物的数据结构, 用图 4 形象地表达了它们各自的结构特点。



4(a) 点状地物的结构



4(b) 弧段(线)的结构



4(c) 面状地物的结构



4(d) 复杂地物的结构

图 4 四类地物的数据结构示意图

## 4 结 束 语

本文提出的GIS中的一体化数据结构是建立在线性四叉树编码基础上的,既有矢量特点又有栅格性质。一方面它保持了较好的精度,而且它可以保持矢量结构的各种拓朴性质,以致于可进行诸如相连关系、相邻关系以及最短距离等拓朴查询。另一方面,它的栅格性质也有助于各种拓朴关系的建立。作者已经发展了一种自动建立拓朴关系的方法,它兼顾使用了这种数据结构的矢量和栅格性质,自动寻找公共结点和自动组织多边形,其速度明显快于单纯基于矢量的系统。<sup>[7]</sup>

这种数据结构的操作是方便的。它可应用类似于操作矢量文件的方法进行增、删、改等数据文件操作,每次修改数据并不会引起数据文件整体结构的破坏。

本文所介绍的矢量栅格一体化数据结构,是整体空间数据库研究的一部分内容。笔者开发了一系列基于这种数据结构的新的算法,如自动建立拓朴关系和面向目标的数据模型,矢量数据到线性四叉树的直接转换以及线性四叉树到矢量数据的直接转换等等,<sup>[7]</sup>从而有可能形成基于这种思想的数据结构、数据模型和数据库体系以及各种操作运算的方法体系,<sup>[7]</sup>同时为建立一个容属性、矢量、栅格、DEM和遥感数据为一体的多功能地理信息系统提供技术基础。

## 参 考 文 献

- [1] Manfred Ehlers Geoffrey Edwards, Yuan Bédard Integration of Remote Sensing with Geographic Information Systems: A Necessary Evolution. *Photogrammetry Engineering and Remote Sensing* Nov. 1989
- [2] Kaehler M R, Theissing U Cartographic Raster Archives-the First Step of An Hybrid Geographic Information System Concept Proce of Conference of EGIS'90, The Netherlands 1990
- [3] Molenaar M and Fritsch D Combined Data Structure for Vector and Raster Representation in Geographic Information Systems Proce of Symp of Comm III of ISPRS Wuhan China 1990
- [4] 吴健康 智能地理信息系统KGIS的设计 测绘出版社, 1988
- [5] 龚健雅 一种基于自然数的线性四叉树编码 测绘学报, No 2 1992
- [6] Yong Hongguang Geometry and Topology of Quadtree Decomposition for GIS Transformation from Raster to Vector Data Proce of EGIS'90, The Netherlands 1990
- [7] 龚健雅 GIS中矢量栅格一体化数据结构与面向目标数据模型的研究 武汉测绘科技大学博士论文, 1991
- [8] Samet H The Quadtree and Related Hierarchical Data Structure *Computing Surveying* No 2, 1984
- [9] Mark D and Lauzon J A Review of Quadtree Based Strategies for Interfacing Coverage Data with Digital Elevation Models in Grid Form *INT J Geographical Information Systems* No 1, 1989

## AN UNIFIED DATA STRUCTURE BASED ON LINEAR QUADTREES

Gong JianYa

*(Wuhan Technical University of Surveying and Mapping)*

### Abstract

A data structure plays an outstanding role in GIS. It is directly related to the function and efficiency of a GIS. A new generation GIS should be capable of integrating various data of GIS, DEM and RS. To serve this end, this paper proposes an unified data structure which has the properties peculiar to both vector and raster. This structure is based on linear quadtree encoding for increasing the geometric accuracy of raster representation and speeding database indexing use is made of multi-grid technique. It can replace conventional vector and raster data structure and has the advantages of maintaining better accuracy, compressing storage space, representing topological and spatial relations among various features and supporting object-oriented data models. It can serve as a base of future GIS.

**Key words** GIS Linear quadtree encoding Unified data structure Object-Oriented data models

(责任编辑: 俞雅珍)

(Continued from page 258)

boundary condition is slope-differential, solving is inconvenient. This paper sufficiently makes use of advantages of Stokes' problem and Molodensky' s problem, and proposes a new method determining shape of the earth and outer gravity field which is based on computing terms of high order in Taylor' s expansion. This method not only has theoretical precision but also has property to compute conveniently in actual application.

**Key words** Nonlinear Molodensky' s problem High order approach

(责任编辑: 徐 侃)